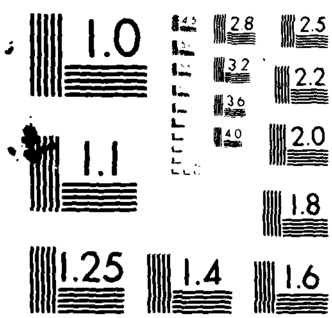
AD-A110 209 CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER F/G 12/1
THE CHARACTERIZATION OF STRICTLY PROPER SCORING RULES IN DECISI--ETC(U)
OCT 81 E HAIM AFOSR-81-0122
UNCLASSIFIED ORC-81-22 NL

END
DATE
FILMED
2 82
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II (12)

ORC 81-22
OCTOBER 1981

THE CHARACTERIZATION OF STRICTLY PROPER SCORING RULES IN DECISION MAKING

by
EDUARDO HAIM

OPERATIONS
RESEARCH
CENTER

8

UNIVERSITY OF CALIFORNIA · BERKELEY

270750

THE CHARACTERIZATION OF STRICTLY PROPER

SCORING RULES IN DECISION MAKING

by

Eduardo Haim
Operations Research Center
University of California, Berkeley

A

OCTOBER 1981 ORC 81-22

This research was supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-81-0122 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC-81-22	2. GOVT ACCESSION NO. AD-A110209	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE CHARACTERIZATION OF STRICTLY PROPER SCORING RULES IN DECISION MAKING		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Eduardo Haim		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0122
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332		12. REPORT DATE October 1981
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Scoring Rules
Forecasting
Decision Theory
Utility Theory
Convexity

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

ABSTRACT

A strictly proper scoring rule or admissible probability measurement procedure (APMP) is a mathematical device that allows a decision-maker to "score" probabilistic forecasts made by experts once the outcome is known. The expected score, as seen by the forecaster, should be maximized when the forecaster states his true beliefs so that he will be encouraged to do so. Applications to subjective probability assessment will be given.

THE CHARACTERIZATION OF STRICTLY PROPER SCORING RULES IN DECISION MAKING

by Eduardo Haim

1. INTRODUCTION

A convenient way to express beliefs on future outcomes is to give a forecast weighted by a probability distribution. For example, in weather forecasting, the forecaster attaches probabilities to the events that snow, rain or sun will occur. These probabilistic forecasts which describe the beliefs of an expert may be used by decision makers in their analysis of a problem. It is, therefore, important to be able to "score" the expert based on past performance in order to attach a "degree of credibility" to the forecast. It is also important to get a forecast which is representative of the expert's "true" beliefs. The expert should not feel that by stating a hedged forecast he will get a higher score.

A strictly proper scoring rule or admissible probability measurement procedure (APMP) is a mathematical device that allows the user (or decision maker) to score the expert's forecast in such a way that the expert will be "encouraged" to state his "true" beliefs. These concepts will be defined more rigorously in Section 2. In Section 3, the theory that will help in the analysis and construction of scoring rules is presented. In Section 4, the results of the previous section are used to construct a strictly proper scoring rule for a given model. The summary, conclusions and topics for further research are in Section 5.

This paper deals exclusively with forecasts weighted by discrete probability distributions, the continuous case will be treated in another paper. The main reference to this work is a 1971 paper by L.J. Savage [3] on the elicitation of personal probabilities. Application of scoring rules to weather forecasting and some desirable properties of scoring rules are presented in References [4] and [5].

An important aspect which is often overlooked in technical papers is stating the reasons why a given technique is useful. Some applications of scoring rules are:

- Elicitation of subjective probabilities;
- Evaluation of expert opinion;
- Use of past performance as represented by the scores to weigh diverging opinions among experts;
- Multiple choice tests could allow for other than absolute forecasts and use scoring rules to evaluate the results.

The reader is referred to the paper by Savage 1971 [3] for other uses of scoring rules.

2. DEFINITIONS AND EXAMPLES

Expert opinion is often given in the form of a probability vector. For example, a weather forecaster can give a forecast of snow, rain or sunshine as a vector (r_1, r_2, r_3) such that

$$r_1 + r_2 + r_3 = 1;$$
$$0 \leqslant r_i \leqslant 1, \; i = 1,2,3 \; ,$$

where

r_1 is the probability of snow

r_2 is the probability of rain

r_3 is the probability of sunshine.

Once the outcome is known, the users of the forecast would like to score the weather forecaster depending on "how close" to the outcome his forecast was.

Definition 1:

Given n possible events, $n \in Z_+$ (positive integers), a forecast may be represented as a probability vector r such that

$$r = (r_1, \ldots, r_n), \; r \in R^n$$
$$0 \leqslant r_i \leqslant 1, \quad i = 1, \ldots, n \; .$$
$$\sum_{i=1}^{n} r_i = 1 \; .$$

Let the outcome be k, $1 \leqslant k \leqslant n$, $k \in Z_+$. Let the forecaster's score be given by the real value $S_k(r)$. We call the real-valued mapping S from (r,k) to $S_k(r)$ a *scoring rule*. A score may be thought of as a reward or as a penalty. In this paper, we shall think of it as a reward so the forecaster wishes to increase his score.

Let the vector $p = (p_1, \ldots, p_n)$, $p \in R^n$ denote the forecaster's true belief (judgment), where

$$0 \leqslant p_i \leqslant 1, \; i = 1, \ldots, n, \text{ and } \sum_{i=1}^{n} p_i = 1 \; .$$

Then p_i represents the probability the forecaster attaches to the occurrence of outcome i.

The stated probability forecast r may not be equal to the forecaster's true belief represented by p. However, the users of the forecast would like to use scoring rules that encourage the forecaster to make his forecast correspond exactly with his true belief. That is, to make $r=p$.

Definition 2:

Let $S_k(r)$ denote the score assigned by the scoring rule S when outcome k occurs and r is the stated probability forecast. Let $ES(r,p)$ be the forecaster's expected score where the vector p is defined as above and

$$ES(r,p) = \sum_{k=1}^{n} p_k S_k(r) \ .$$

Then, the scoring rule S is said to be *strictly proper* if

$$ES(p,p) \geqslant ES(r,p) \quad \forall\, p, \textit{ with equality} \text{ if and } \textit{only} \text{ if } r = p \ .$$

In other words, S is a *strictly proper* scoring rule if and only if the forecaster's expected score is maximized uniquely when he sets $r=p$.

S is a *proper* scoring rule if p maximizes ES but not uniquely.

Example 1: Brier Scoring Rule

Let,

$r \in R^n$, the stated probability forecast vector.

$p \in R^n$, the forecaster's true belief probability vector.

Let n be number of possible outcomes and

$k \in Z_+, 1 \leqslant k \leqslant n$, the outcome.

Define

$$S_k(r) = -(r - e^k)^T I (r - e^k)$$

where

$$e^k = \begin{cases} 1 & \textit{in position } k \\ 0 & \textit{elsewhere} \end{cases}$$

e^k is a vector of dimension n which indicates that outcome k has occurred and I is the identity matrix of dimension n.

We wish to show that the Brier Scoring Rule is strictly proper. Note that

$ES(p,p) - ES(r,p) = (r-p)^T I (r-p) \geqslant 0 \quad \forall p$, with equality if and only if r=p since I is positive definite. Then,

$$ES(p,p) \geqslant ES(r,p) \quad \forall p, \textit{ with equality if and only if } r = p .$$

Therefore, the Brier Scoring Rule is strictly proper.

Note that the proof would be the same if I were replaced by any positive definite matrix of dimension n. In this way the family of quadratic scoring rules is generated.

Example 2:

Let $r \in R^n$ be the stated probability forecast vector.

Let $p \in R^n$ be the forecaster's true belief probability vector.

Let n be the number of possible outcomes and $k \in Z_+$, $1 \leqslant k \leqslant n$, the outcome.

Define $T_k(r) = -e^T|r - e^k|$, a scoring rule. e^k is the same vector as in Example 1 and e is a column vector of ones. Both have dimension n. Suppose $n=2$, then

$$\begin{aligned} \textit{Expected score} &= ET(r,p) \\ &= -[p_1(|r_1 - 1| + |r_2 - 0|) + p_2(|r_1 - 0| + |r_2 - 1|)] \\ &= -[2p_1r_2 + 2p_2r_1] . \end{aligned}$$

$ET(r,p)$ is maximized by

$$(r_1,r_2) = \begin{cases} (1,0) & \text{if } p_1 \geqslant p_2 \\ (0,1) & \text{if } p_1 < p_2 \end{cases}.$$

So the scoring rule T is *not* proper.

When a scoring rule is not as simple as the ones presented in the above examples, it is harder to prove that it is strictly proper through the expected score definition of the property. In Section 3, a theorem is presented that will make it easier to demonstrate that a scoring rule is strictly proper. The theorem can also be used to construct strictly proper scoring rules. In Section 4, the problem of constructing scoring rules which better represent the interests of the users of the forecast will be discussed.

2.1. The Utility Function of the Expert (Forecaster)

From a decision analysis point of view, the expert forecaster should be maximizing his expected utility when he states his forecast. It is technically incorrect to say that the forecaster wants to maximize his expected score.

If the utility function of the forecaster is a real valued function f such that

$$(I) \qquad \begin{cases} \sum_{k=1}^{n} S_k(r) p_k \leq \sum_{k=1}^{n} S_k(p) p_k \quad \forall p, \textit{ with equality iff } r=p \\ \text{if and } \textit{only} \text{ if} \\ \sum_{k=1}^{n} f(S_k(r)) p_k \leq \sum_{k=1}^{n} f(S_k(p)) p_k \quad \forall p, \textit{ with equality iff } r=p \end{cases}$$

then maximizing expected score is equivalent to maximizing expected utility.

Let

$$g : R \to R \textit{ be such that}$$
$$g(x) = ax + b \quad a \in R, a > 0, b \in R .$$

It is easy to show that condition (I) holds for such a function g.

Necessary and sufficient conditions on f for (I) to hold are very restrictive [6]. However, since (I) holds for the linear increasing function g, it will be assumed that the utility function of the forecaster is piecewise linear and increasing in the score. This assumption is not too restrictive since any smooth function can be arbitrarily approximated by piecewise linear functions and the utility function of the expert should be increasing in the score. For relatively small changes in the value of the score, the utility function can be assumed locally linear.

Based on the above, maximizing expected score will be assumed to be equivalent to maximizing expected utility.

2.2. Perfect Forecasts

Definition 3:

A forecast that assigns probability one to one of the possible outcomes and probability zero to all the others is defined as an *absolute forecast.* The vector r such that

$$r_i = \begin{cases} 1 & \text{if } i = k \\ 0 & \textit{otherwise} \end{cases}$$

is an absolute forecast of outcome k.

Definition 4:

A *perfect forecast* is an absolute forecast of the actual outcome.

Theorem 1:

If a scoring rule is (strictly) proper and the outcome is k, the score for outcome k is maximized (uniquely) by a perfect forecast of k.

Proof:

Let S be a strictly proper scoring rule, then

$$\sum_{i=1}^{n} S_i(r)p_i \leq \sum_{i=1}^{n} S_i(p)p_i \quad \forall p, \textit{ with equality} \text{ if and } \textit{only} \text{ if } r = p \ ,$$

or

$$\sum_{i=1}^{n} (S_i(p) - S_i(r))p_i \geq 0 \quad \forall p, \textit{ with equality} \text{ if and } \textit{only} \text{ if } r = p \ .$$

Let p be a perfect forecast of outcome k.

$$p_i = \begin{cases} 1 & \text{if } i = k \\ 0 & \textit{otherwise} \end{cases}$$

then

$$(S_k(p) - S_k(r)) \cdot 1 \geqslant 0, \textit{ with equality} \text{ if and } \textit{only} \text{ if } r = p$$

$$S_k(p) \geqslant S_k(r), \textit{ with equality} \text{ if and } \textit{only} \text{ if } r = p \;. \blacksquare$$

3. NECESSARY AND SUFFICIENT CONDITIONS FOR A SCORING RULE TO BE STRICTLY PROPER

Definition 5:

A *subset* C *of* R^n is said to be *convex* if $(1-\lambda)x+\lambda y \in C$ where $x \in C$, $y \in C$ and $0<\lambda<1$.

Definition 6:

A *real-valued function* f on a convex set C is said to be *(strictly) convex* on C if,

$$f((1-\lambda)x_1+\lambda x_2)\ (<) \leqslant (1-\lambda)f(x_1)+\lambda f(x_2),\ 0<\lambda<1$$

for any two different points x_1 and x_2 in C.

Definition 7:

A vector x^* is said to be a *subgradient* of a convex function f at a point x if,

$$f(y) \geqslant f(x) + \langle (y-x), x^* \rangle \quad \forall y$$

where $\langle \cdot , \cdot \rangle$ is the usual inner product. This condition is referred to as the subgradient *inequality*[1]. If f is strictly convex, equality occurs if and only if $x=y$.

Theorem 2:

A scoring rule S is strictly proper if and only if

$$S_k(r) = G(r) - \langle r, G^*(r) \rangle + G_k^*(r) + a_k$$

where r is the vector of probability forecasts of dimension n, k is the outcome, $1 \leqslant k \leqslant n$.

$$G^*(r) = (G_1^*(r), \ldots, G_k^*(r), \ldots, G_n^*(r))$$

is a subgradient at the point r of the strictly convex function G, and a_k is a constant that may depend on the outcome k but not on the vector of probability forecasts r.

(1) The subgradient inequality has a simple geometric meaning when f is finite at x: it says that the graph of the affine function $h(y)=f(x)+\langle (y-x), x^* \rangle$ is a non-vertical supporting hyperplane to the convex set epi f at the point $(x, f(x))$. See Rockafellar 1970, Section 23 [Reference 2].

Proof:

Suppose

$$S_k(r) = G(r) - \langle r, G^*(r) \rangle + G_k^*(r) + a_k$$

where $G^*(r)$ is a subgradient at the point r of the strictly convex function G and a_k is a constant that depends on k but not on r.

We wish to show that S is a strictly proper scoring rule. Let $p=(p_1,...,p_n)$ be the forecaster's true judgment and $a=(a_1,...,a_n)$, then

$$\begin{aligned} ES(p,p) - ES(r,p) = G(p) - \langle p, G^*(p) \rangle + \langle p, G^*(p) \rangle + \langle p, a \rangle \\ - [G(r) - \langle r, G^*(r) \rangle + \langle p, G^*(r) \rangle + \langle p, a \rangle] \end{aligned}$$

therefore,

$$ES(p,p) - ES(r,p) = G(p) - G(r) - \langle (p-r), G^*(r) \rangle .$$

Since $G^*(r)$ is a subgradient at the point r of the strictly convex function G,

$$G(p) \geqslant G(r) + \langle (p-r), G^*(r) \rangle \quad \forall p, \textit{ with equality if and only if } r = p .$$

Therefore,

$$ES(p,p) \geqslant ES(r,p) \quad \forall p, \textit{ with equality if and only if } r = p .$$

So S is a strictly proper scoring rule.

Conversely, suppose S is a strictly proper scoring rule, that is,

$$ES(p,p) \geqslant ES(r,p) \quad \forall p, \textit{ with equality if and only if } r = p .$$

Let

$$G(p) = ES(p,p) = \langle p, S(p) \rangle ,$$
$$G_k^*(p) = S_k(p) ,$$

and

$$a_k = 0 ,$$

then

$$S_k(p) = G(p) - \langle p, G^*(p) \rangle + G_k^*(p) + a_k .$$

To show that G is strictly convex and $G^*(p)$ is a subgradient at the point p of G, note that

$$G(p) \geqslant ES(r,p) = \langle p, S(r) \rangle \quad \forall p, \textit{ with equality } \text{if and } \textit{only} \text{ if } r = p$$

since S is strictly proper.

For each fixed r, the function $ES(r,p)$, which is linear in p, lies strictly below the function G, except at r, where the linear function and G have the common value $ES(r,r)=G(r)$. In short, $ES(r,p)$ is, for each r, a linear function of support of G at the point r and only there. Therefore, G is a strictly convex function of p. (This follows from [2] Rockafellar (1970), Theorem 5.5, page 35.)

Now consider,

$$\begin{aligned} & G(p) - G(r) - \langle (p-r), G^*(r) \rangle \\ &= \langle p, S(p) \rangle - \langle r, S(r) \rangle - \langle p, S(r) \rangle + \langle r, S(r) \rangle \\ &= \langle p, S(p) \rangle - \langle p, S(r) \rangle = ES(p,p) - ES(r,p) \geqslant 0 \end{aligned}$$

$\forall p$, with equality if and only if $r=p$, since S is a strictly proper scoring rule. So,

$$G(p) \geqslant G(r) + \langle (p-r), G^*(r) \rangle \quad \forall p, \textit{ with equality } \text{if and } \textit{only} \text{ if } r = p \,.$$

Therefore, $G^*(r)$ is a subgradient at the point r of the strictly convex function G.∎

Corollary 1:

A scoring rule S is strictly proper if

$$S_k(r) = G(r) - \langle r, \nabla G(r) \rangle + G_k'(r) + a_k$$

where r is the vector of probability forecasts of dimension n and k is the outcome, $1 \leqslant k \leqslant n$.

$$\nabla G(r) = \left(\frac{\partial G(r)}{\partial r_1}, \ldots, \frac{\partial G(r)}{\partial r_n} \right) = (G_1'(r), \ldots, G_k'(r), \ldots, G_n'(r))$$

is the gradient at the point r of the strictly convex, differentiable function G, and a_k is a constant that may depend on the outcome k but not on the vector of probability forecasts r.

Proof:

The gradient is also a subgradient since

$$G(p) \geqslant G(r) + \left\langle (p - r)\,,\, \nabla G(r) \right\rangle \quad \forall\, p, \text{ with equality if and only if } r = p\ .$$

Apply Theorem 2 and the result follows. ■

The results of this section are quite important. They provide us with another way to recognize strictly proper scoring rules, but more important still, they give us a method to construct them. In Section 4, the results of this section are used in the construction of a strictly proper scoring rule.

4. CONSTRUCTION OF STRICTLY PROPER SCORING RULES BY THE USER. AN APPLICATION

In Section 2, the Brier Scoring Rule was presented. It has the desirable property of being strictly proper. However, it does not allow for the user to be represented in the scoring rule. One way the user can be taken into consideration is the inclusion of his utility function in the scoring rule.

In this section, we construct a scoring rule that is a function of the user's utility. The construction is based on Theorem 2 of Section 3 and it illustrates the way in which this theorem can be used to construct strictly proper scoring rules for different situations.

4.1. The Model

Suppose there are n possible outcomes or events $E_1,\dots,E_n$ and that the user of the forecast or decision maker can take one of m possible actions $A_1,\dots,A_m$.

Let u_{ij} be the utility of the user when the action he took was A_i and the outcome is E_j. Then there is a matrix U of utilities

$$U = \begin{array}{c} \\ A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_m \end{array} \begin{array}{c} \begin{array}{ccccc} E_1 & \cdot & \cdot & \cdot & E_n \end{array} \\ \begin{bmatrix} u_{11} & \cdot & \cdot & \cdot & u_{1n} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & & u_{ij} & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ u_{m1} & \cdot & \cdot & \cdot & u_{mn} \end{bmatrix} \end{array}$$

The value of the utility of each point of matrix U is relative to the utility of the other points. Without loss of generality it is assumed that $u_{ij}>0 \ \forall\, i,j$.

The forecaster attaches a probability to the occurrence of each event; $r=(r_1,...,r_n)$; $\sum_{i=1}^{n} r_i=1$; $r_i \geqslant 0, i=1,...,n$, where r is the vector of probability forecasts.

4.2. Construction of a Strictly Proper Scoring Rule

Given the above model, a strictly proper scoring rule is constructed based on Theorem 2 of Section 3.

(i) Let $G(r) = e^{\max_i \langle r, U_i \rangle}$

where U_i is row i of matrix U.

(ii) Let $G^*(r) = \nabla G(r) =$ *gradient of G at r* .

$$G_k^*(r) = G_k'(r) = \frac{\partial G(r)}{\partial r_k} \quad k = 1,...,n$$

for all $r \in C$ for which G is differentiable where the set

$$C = \left\{ r \mid r \in R^n;\ \sum_{i=1}^{n} r_i = 1;\ r_i \geqslant 0,\ i = 1, \ldots, n \right\} \textit{ is convex} .$$

If $r \in C$ and G is not differentiable at the point r, $G^*(r)$ is defined as the right (side directional) derivative of G at the point r.

$$G^*(r) = G'_+(r) = G'(r,1) = \lim_{\alpha \downarrow 0} \frac{G(r+\alpha) - G(r)}{\alpha} .$$

Now $G^*(r)$ is well defined since G is strictly convex [see proof below] and, therefore, the directional derivative exists for all $r \in C$ [Reference 2]. Also, $G'_+(r) = \nabla G(r)$ if G is differentiable at r.

(iii) Let a_k be a constant that depends on the kth outcome.

Claim 1:

$G(r) = e^{\max_i \langle r, U_i \rangle}$ is strictly convex.

Proof:

$$G(r) = e^{\max_i \langle r, U_i \rangle} = \max_i e^{\langle r, U_i \rangle}$$

since e^x is increasing in x.

Define

$$\overline{G}(r,i) = e^{\langle r, U_i \rangle} \quad i = 1, \ldots, m \;.$$

The Hessian H of $\overline{G}(r,i)$ with respect to r follows:

$$H = e^{\langle r, U_i \rangle} \begin{bmatrix} u_{i1}^2 & \cdot & \cdot & \cdot & \cdot & u_{i1}u_{in} \\ \cdot & u_{i2}^2 & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ & & & \cdot & & \\ \cdot & & & & \cdot & \cdot \\ u_{i1}u_{in} & & \cdot & \cdot & \cdot & u_{in}^2 \end{bmatrix}$$

We need to show that H is a positive definite matrix.

Let y be an arbitrary vector of dimension n, then

$$y^T H y = e^{\langle r, U_i \rangle} \cdot (y_1 u_{i1} + \cdots + y_n u_{in})^2 \;.$$

Recall that, by assumption, $u_{ij} > 0$ for all i,j. Therefore, $y^T H y > 0$ for all $y \neq 0$, so H is positive definite. This implies that $\overline{G}(r,i)$ is strictly convex. Since

$$G(r) = \max_i \overline{G}(r,i) \;,$$

it follows that G is a strictly convex function of r. ∎

$G^*(r)$ is a subgradient at the point r of the strictly convex function G by definition. Before using Theorem 2 to construct a strictly proper scoring rule, the following set is defined,

$$D_i = \{r \mid \langle r, U_i \rangle \geq \langle r, U_j \rangle \quad \forall j \neq i\} \qquad i = 1, \ldots, n .$$

Claim 2:

D_i is a convex set, $i=1,\ldots,n$.

Proof:

$$r,t \in D_i \textit{ implies that } \begin{cases} \langle r, U_i \rangle \geq \langle r, U_j \rangle & \forall j \neq i \\ \langle t, U_i \rangle \geq \langle t, U_j \rangle & \forall j \neq i \end{cases} .$$

Then,

$$\lambda \langle r, U_i \rangle + (1-\lambda)\langle t, U_i \rangle \geq \lambda \langle r, U_j \rangle + (1-\lambda)\langle t, U_j \rangle$$
$$\forall j \neq i \text{ and } 0 < \lambda < 1 .$$

So,

$$\langle (\lambda r + (1-\lambda)t),\, U_i \rangle \geq \langle (\lambda r + (1-\lambda)t),\, U_j \rangle$$
$$\forall j \neq i \text{ and } 0 < \lambda < 1 .$$

Therefore, D_i is a convex set.■

Now, using the sets D_i, $i=1,\ldots,n$

$$G(r) = e^{\langle r, U_i \rangle} \qquad r \in D_i$$
$$G_k^*(r) = u_{ik} e^{\langle r, U_i \rangle} \qquad r \in D_i \,;\ [r, r+\epsilon) \subset D_i$$

where $\epsilon=(\epsilon_1,\ldots,\epsilon_n)$ for some ϵ such that $\epsilon_k>0$. Let the scoring rule S be as follows,

$$S_k(r) = G(r) - \langle r, G^*(r) \rangle + G_k^*(r) + a_k$$
$$S_k(r) = e^{\langle r, U_i \rangle} - \langle r, U_i e^{\langle r, U_i \rangle} \rangle + u_{ik} e^{\langle r, U_i \rangle} + a_k \qquad r \in D_i;\ [r, r+\epsilon) \subset D_i .$$

So,

$$S_k(r) = e^{\langle r, U_i \rangle}(1 - \langle r, U_i \rangle + u_{ik}) + a_k \qquad r \in D_i;\ [r, r+\epsilon) \subset D_i .$$

S is well suited for the model presented in Section 4.1. The score that the forecaster gets is a function of the expected utility of the user when he takes action A_i and also of the utility of the specific outcome. The term a_k depends only on the outcome k and can be disregarded

while interpreting the score.

$\langle r, U_i \rangle$ is the expected utility of the user given the forecast $r \in D_i$. u_{ik} is the utility obtained from having taken action A_i when the outcome is E_k. u_{ik} is the term that most clearly improves the forecaster's score.

By construction, S is a strictly proper scoring rule [Theorem 2]. The score for outcome k is maximized uniquely by a perfect forecast of k [Theorem 1].

For comparison, suppose

$$G(r) = \langle r, U_i \rangle \qquad r \in D_i$$
$$G_k^*(r) = u_{ik} \qquad r \in D_i \,;\, [r, r+\epsilon) \subset D_i \quad .$$

Note that $G(r)$ is now convex (not strictly convex.) Following the same procedure as before we get,

$$S_k(r) = u_{ik} + a_k \qquad r \in D_i \,;\, [r, r+\epsilon) \subset D_i$$

which is a proper scoring rule (not a strictly proper scoring rule).

5. SUMMARY AND CONCLUSIONS

A strictly proper scoring rule was constructed which takes into consideration the utility function of the user of the forecast. The scoring rule was presented as an application of Theorem 2 in Section 3 which gives the necessary and sufficient conditions for a scoring rule to be strictly proper.

Suppose a model of a given situation is presented. By finding a strictly convex function that accounts for some part of the behavior of the model, a strictly proper scoring rule can be constructed following the procedure of Section 4. More applications of this methodology will be presented in a forthcoming paper. Also, the case of a continuous density function forecast (as opposed to a discrete probability forecast) will be studied separately.

Different scoring rules are needed because forecasters are not ideal subjects. In theory, any strictly proper scoring rule should encourage the forecaster to state his true beliefs. However, this does not happen in practice and the need arises for scoring rules adapted to each situation. Also, the forecaster being subjected to evaluation by scoring rules should have a minimum knowledge of probabilistic concepts and of utility theory in order to realize that his objective is to maximize his expected score.

REFERENCES

[1] Lindley, Dennis V., "Scoring Rules and the Inevitability of Probability," ORC 81-1, Operations Research Center, University of California, Berkeley, (1981).

[2] Rockafellar, R. Tyrrell, CONVEX ANALYSIS, Princeton University Press, Princeton, New Jersey, (1970).

[3] Savage, Leonard J., "Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association, Vol. 66, No. 336, (December 1971).

[4] von Holstein, Carl-Axel S. Stael, "A Family of Strictly Proper Scoring Rules Which Are Sensitive to Distance," Journal of Applied Meteorology, Vol. 9, (June 1970).

[5] von Holstein, Carl-Axel S. Stael and Allan H. Murphy, "The Family of Quadratic Scoring Rules," Monthly Weather Review, Vol. 106, No. 7, (July 1978).

[6] Mak, King-Tim, oral communication, forthcoming paper, (1981).

DAT
FILM